

**Lieff
Cabraser
Heimann &
Bernstein**
Attorneys at Law

SUSMAN GODFREY L.L.P.
A REGISTERED LIMITED LIABILITY PARTNERSHIP



November 19, 2024

VIA ECF

Hon. Ona T. Wang
Daniel Patrick Moynihan
United States Courthouse
500 Pearl St.
New York, NY 10007

RE: *Authors Guild v. OpenAI Inc.*, (No. 1:23-cv-08292-SHS)
Alter v. OpenAI Inc., (No. 1:23-cv-10211-SHS)

Dear Judge Wang:

Pursuant to Rule II(b) of Your Honor's Individual Practices, Plaintiffs seek a conference regarding OpenAI's refusal to produce documents and training data related to certain OpenAI models, including GPT-5, Orion, commercial Open AI models, research models, and models used to power Microsoft's commercial AI products ("the Additional Models").¹ OpenAI seeks to limit discovery to only those models that power ChatGPT and a few others, but there is no basis to limit discovery on the Additional Models. Documents and training data related to the Additional Models are discoverable for two reasons: 1) the complaint alleges that OpenAI's use of Class Works to train these Additional Models is copyright infringement (*i.e.*, it is not limited to models that power ChatGPT) and 2) OpenAI has not demonstrated that this discovery is burdensome.

OpenAI should also be compelled to provide full responses to Plaintiffs' Interrogatories about the Additional Models. Interrogatory No. 10 seeks identification of every version of every LLM, Generative AI system, AI model, and API Product that is or was in OpenAI's possession, custody, or control (including gpt-4-base, a research-only model). Ex. 4 at 4-5. Interrogatory No. 11 seeks identification of any models listed in OpenAI's response to Interrogatory No. 10 that have been used in any version of ChatGPT or any other OpenAI product or service. *Id.*

The parties have met and conferred about OpenAI's Additional Models and were not able to reach agreement. The parties first conferred about the Additional Models in July. OpenAI refused to fully identify all Additional Models in response to Interrogatory No. 10 because doing so would be "unduly burdensome and disproportionate given the lack of relevance." Ex. 5 at 1. Class Plaintiffs later learned that OpenAI agreed to give News Plaintiffs discovery on three Additional Models—GPT, GPT-2, and GPT-3. On October 11, Class Plaintiffs re-urged their requests and asked OpenAI to give Class Plaintiffs similar discovery on GPT, GPT-2, and GPT-3.

¹ Documents and training data related to the Additional Models are responsive to Plaintiffs' RFP Nos. 6, 10, 17, 41, 42, and 43, among others. *See* Exs. 1-3.

November 18, 2024

Page 2

Ex. 6. The parties conferred on October 16. On October 18, Plaintiffs requested “further discussions,” and followed up on October 25. OpenAI agreed to give Class Plaintiffs discovery on GPT, GPT-2, and GPT-3 and to supplement its interrogatory responses to include GPT, GPT-2, and GPT-3, but did not agree to produce discovery about other Additional Models or to identify other Additional Models in response to Interrogatories Nos. 10 and 11. Ex. 7 at 1.

On November 6, 2024, OpenAI’s counsel said it would be productive for the parties to discuss the Additional Models issue again and said that he would propose a time for a call. Ex. 8 at 1. Nine days passed, and OpenAI failed to even propose a time for another meet and confer. On November 15, 2024, Class Plaintiffs’ counsel informed OpenAI that it would file a motion on the Additional Models given the parties’ lack of progress on these issues. Ex. 9 at 1. Any argument by OpenAI that Class Plaintiffs have failed to exhaust the Court’s conferral requirements is belied by the email records and OpenAI’s own delays. Class Plaintiffs now move to compel OpenAI to fully respond to Interrogatories Nos. 10 and 11, to search for responsive documents about the Additional Models, and to make available the training data for the Additional Models.

Because Plaintiffs allege that all OpenAI models infringe, OpenAI must provide discovery on the Additional Models. Plaintiffs allege that *all* OpenAI’s artificial intelligence models infringe because all are trained on the Class Works. *See, e.g.*, First Consolidated Class Action Compl., Dkt. 69 ¶ 413 (“Plaintiffs and members of the Classes own the registered copyrights in the works that Defendants reproduced and appropriated to train their artificial intelligence models.”); *id.* ¶ 127 (“In short, OpenAI admits it needs and uses ‘large, publicly available datasets that include copyrighted works’—and specifically, ‘high-quality’ copyrighted books—to ‘train’ its LLMs; pirated sources of such ‘training’ data are readily available; and one or more of these sources contain Plaintiffs’ works.”); *id.* ¶¶ 394, 397 (alleging that the Class Works are works “used by Defendants to ‘train’ one or more of Defendants’ large language models.”).

Plaintiffs’ claims are not limited to models powering ChatGPT or the GPT, GPT-2, and GPT-3 models, which is an artificial limit that OpenAI seeks to impose on discovery. If OpenAI used Class Works for training of any of the Additional models, that is copyright infringement at issue in the case, and OpenAI cannot cut off discovery on these issues. *See* Fed. R. Civ. P. 26(b)(1) (“Parties may obtain discovery regarding any matter, not privileged, that is relevant to the claim or defense of any party”). OpenAI may disagree on the merits of Plaintiffs’ claims, but it cannot unilaterally define the scope of discovery by misconstruing the allegations of Plaintiffs’ complaint. *See, e.g., XChange Telecom Corp. v. Sprint Spectrum L.P.*, 2015 WL 773752, at *3 (N.D.N.Y. Feb. 24, 2015) (“No party possess[es] the unilateral ability to dictate the scope of discovery based on their own view of the parties’ respective theories of the case.”) (internal quotation marks omitted). OpenAI should be compelled to fully respond to Interrogatories seeking basic data about the Additional Models, to search for responsive documents related to the Additional Models, and to produce training data for the Additional Models.

OpenAI has not proven that discovery on the Additional Models is unduly burdensome. OpenAI objects to the discovery Plaintiffs seek as unduly burdensome. But OpenAI has not met its burden to demonstrate that discovery on the Additional Models is unduly burdensome. *See, e.g., Citizens Union of City of New York v. Att’y Gen. of New York*, 269 F. Supp. 3d 124, 139 (S.D.N.Y. 2017) (“[T]he party withholding discovery on the grounds of burden . . . bears the burden of proving the discovery is in fact . . . unduly burdensome and/or expensive.”). When Plaintiffs asked

November 18, 2024

Page 3

OpenAI how many Additional Models there are, OpenAI refused to even answer the question, and in particular refused to even identify the Additional Models in the first place. Ex. 9. Because OpenAI has not demonstrated that discovery on the Additional Models is burdensome, any burden argument should be rejected.

If the Court considers OpenAI's burden argument, Plaintiffs are willing to confer with OpenAI on potential ways to alleviate undue burden to OpenAI. Any undue burden associated with documents can likely be resolved via the parties' search term negotiation process. Any undue burden associated with providing training data can likely be resolved via negotiations between the parties.

For the reasons discussed above, the Court should compel OpenAI to provide full discovery on the Additional Models.

Sincerely,

LIEFF CABRASER HEIMANN
& BERNSTEIN LLP

SUSMAN GODFREY LLP

COWAN, DEBAETS,
ABRAHAMS & SHEPPARD
LLP

/s/ Rachel Geman
Rachel Geman

/s/ Rohit Nath
Rohit Nath

/s/ Scott J. Sholder
Scott J. Sholder